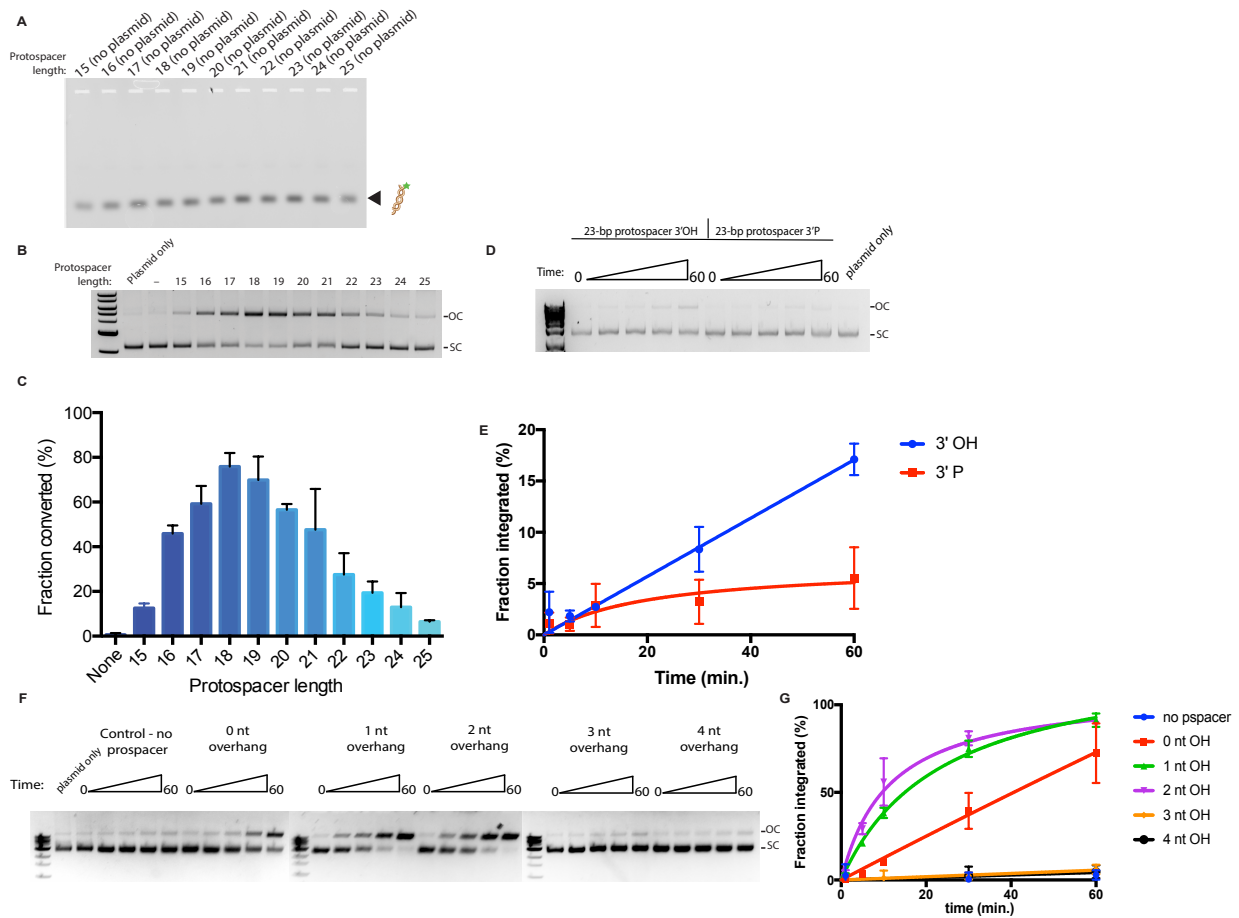# Figure S1

Figure S1. Related to Figure 2. Determination of type V-C Cas1 protospacer preferences

(A) Control integration assay with variable length fluorescent protospacers with no plasmid.

(B) Integration assay with variable length protospacers from 15-bp to 25-bp long.

(C) Quantification of (B) demonstrating the effect of protospacer length on the conversion of supercoiled plasmid to open-circle products by type V-C Cas1. The fraction converted is calculated as the fraction of open-circle products relative to all plasmid. Experiments were carried out in triplicate; the bars represent mean values, with error bars depicting standard deviations.

(D) Time-course integration reactions with a 23-bp protospacer and a 23-bp protospacer with 3' phosphate modifications on both strands.

(E) Quantification of (D). Plotted points represent the mean of three independent experiments. Error bars represent standard deviation.

(F) Time-course integration reactions with protospacers with 0- to 4-nt overhangs.

(G) Quantification of (F). Plotted points represent the mean of three independent experiments. Error bars represent standard deviation.

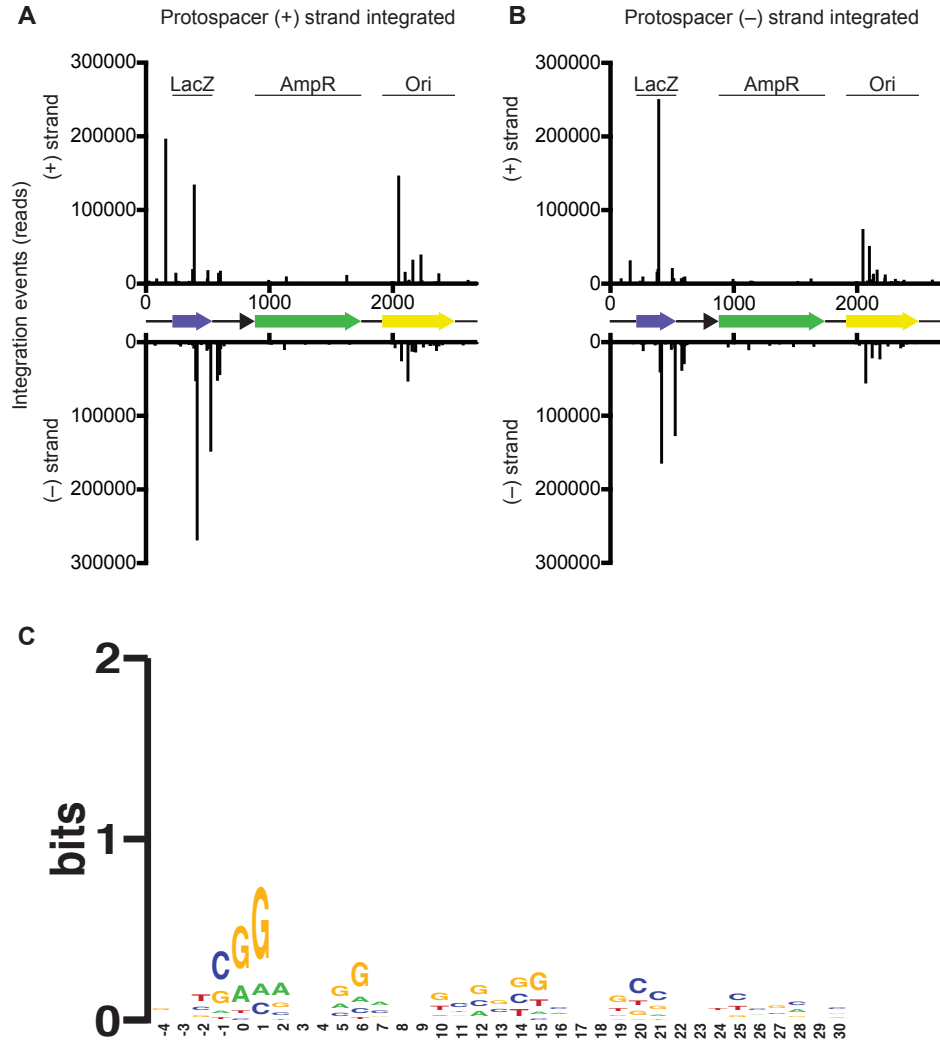See Table S2 for nucleotide sequences.

Figure S2

A Protospacer (+) strand integrated

B Protospacer (−) strand integrated

C

Figure S2. Related to Figure 3. Integration Sites Along pUC19 and WebLogo for Integration Sites in pCRISPR
(A, B) Integration sites along pUC19 for protospacer (+) strand (A) and (−) strand (B).
(C) WebLogo for integration sites in pCRISPR. The first nucleotide following integration is numbered 1, with the upstream nucleotides labeled -4 through 0.

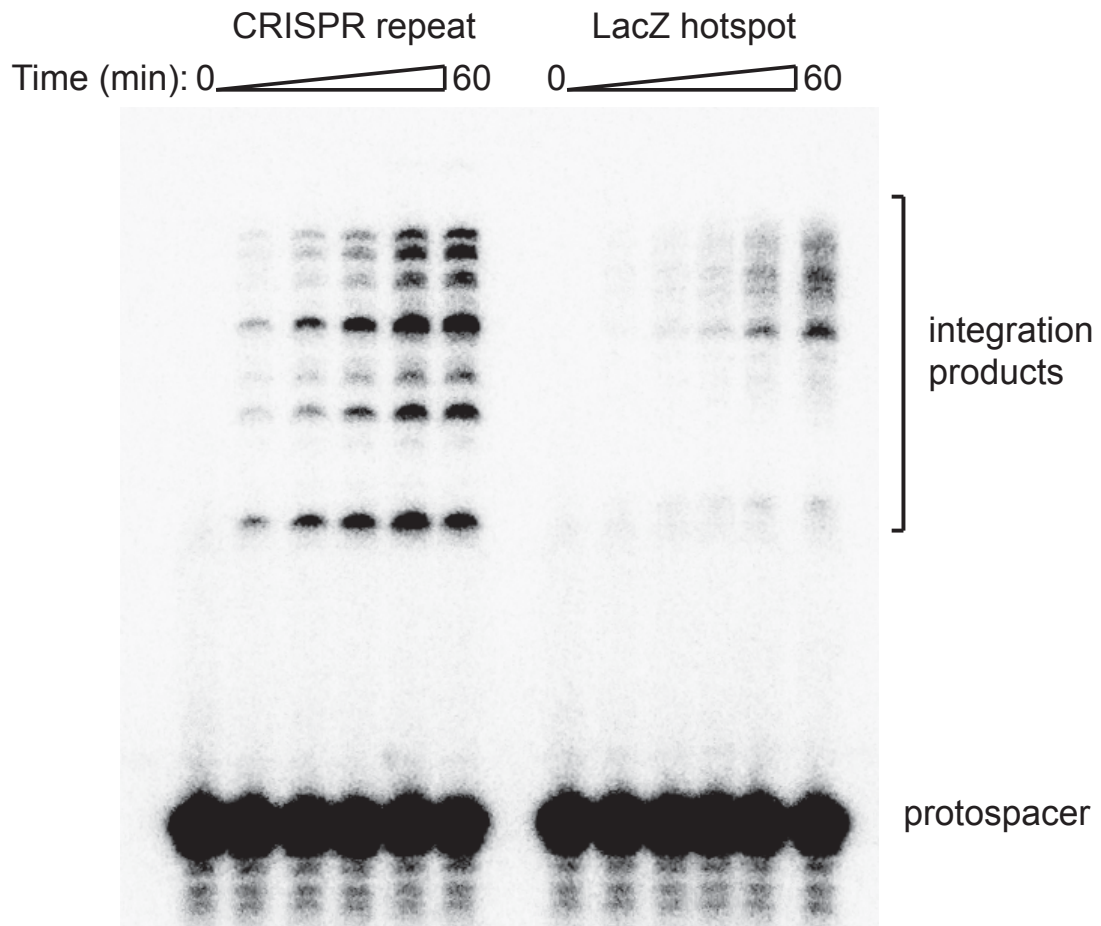See Table S2 for nucleotide sequences.

Figure S3

Figure S3. Related to Figure 3. Integration of Protospacer into CRISPR Repeat and *lacZ* Hotspot
Integration reactions are conducted with radiolabeled protospacer and target plasmids containing a CRISPR repeat or the *lacZ* hotspot identified from high-throughput sequencing. See Table S2 for nucleotide sequences.
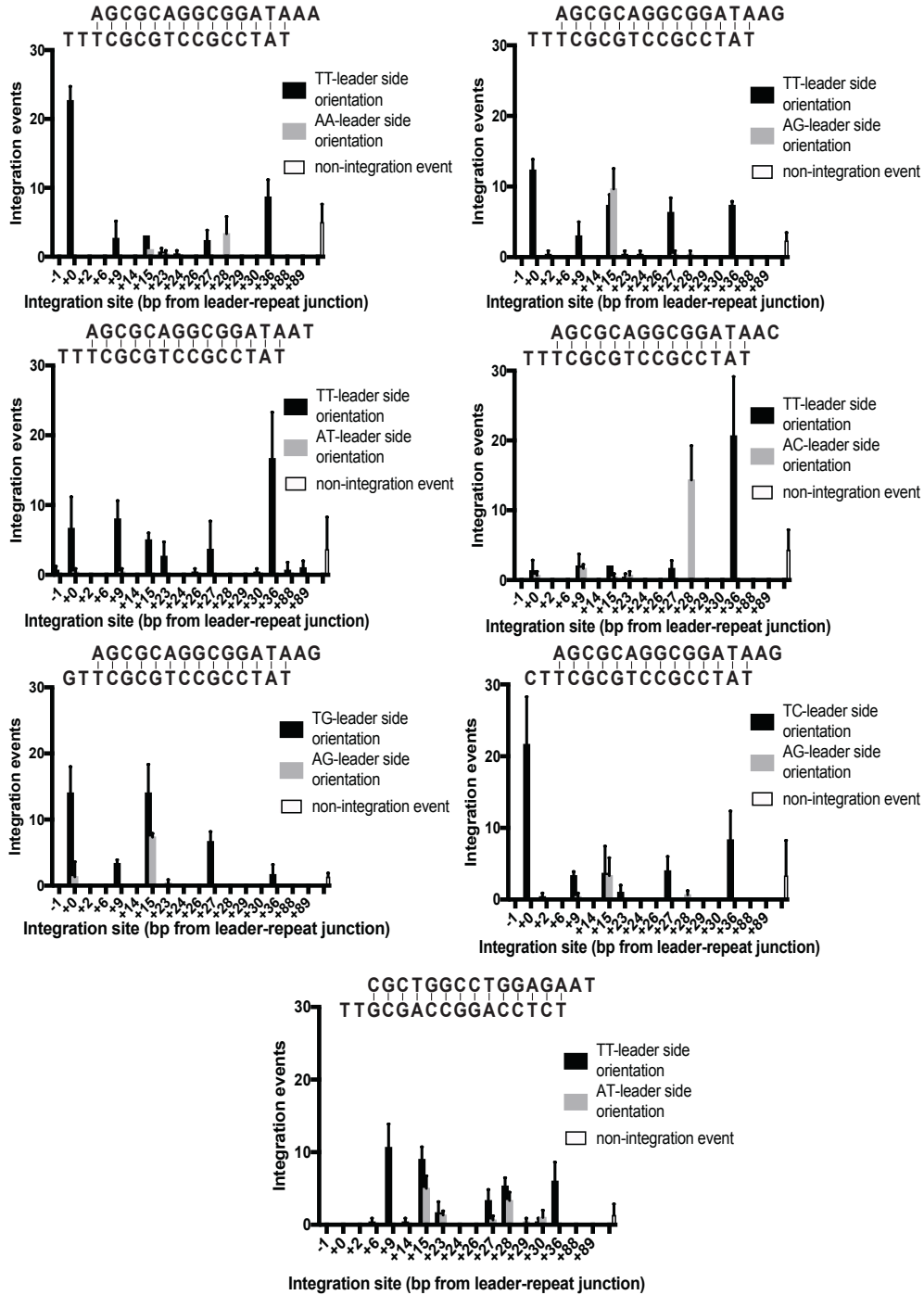
Figure S4

Figure S4. Related to Figure 4. Protospacer Sequence Influences Orientation During Full-Site Integration

Graphs show the number of integration events with specified spacer orientation at each integration site. The protospacer sequence is listed above each graph. The number of non-integration events (as a result of deletions) is also indicated. Mean and standard deviation of three independent replicates are plotted. See Table S2 for nucleotide sequences.

Figure S5
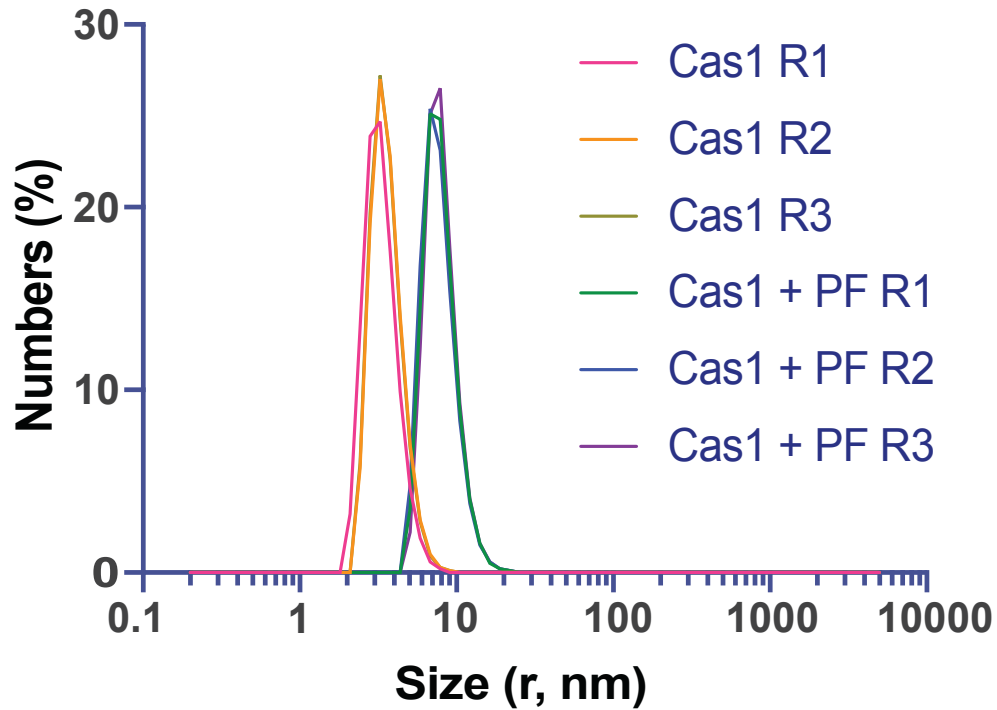
Sample	Mean size (r, nm)	PdI
Cas1 R1	3.8 ± 0.8	0.39
Cas1 R2	3.8 ± 0.8	0.40
Cas1 R3	3.9 ± 0.7	0.45
Cas1 + PF R1	7.7 ± 1.9	0.34
Cas1 + PF R2	8.8 ± 1.9	0.42
Cas1 + PF R3	8.0 ± 2.1	0.43

Figure S5. Related to Figure 5. Dynamic Light Scattering Size Measurements of Apo Cas1 Compared to Cas1 with Added Pseudo-Full-Site Substrate
Measurements were collected from three runs. The mean size with standard deviations and PdI for each sample run are provided below the distribution plots. See Table S2 for nucleotide sequences.

Figure S6

Molecular Mass Legend

A: 29,128 Da B: 31,459 Da C: 33,789 Da

protospacer

pseudo-full-site substrate

leader repeat spacer

D: 43,539 Da E: 87,030 Da

Cas1 monomer *Cas1 dimer*

F: 115,980 Da G: 118,290 Da H: 120,590 Da

Cas1 dimer bound to DNA

I: 179,100 Da

Cas1 tetramer

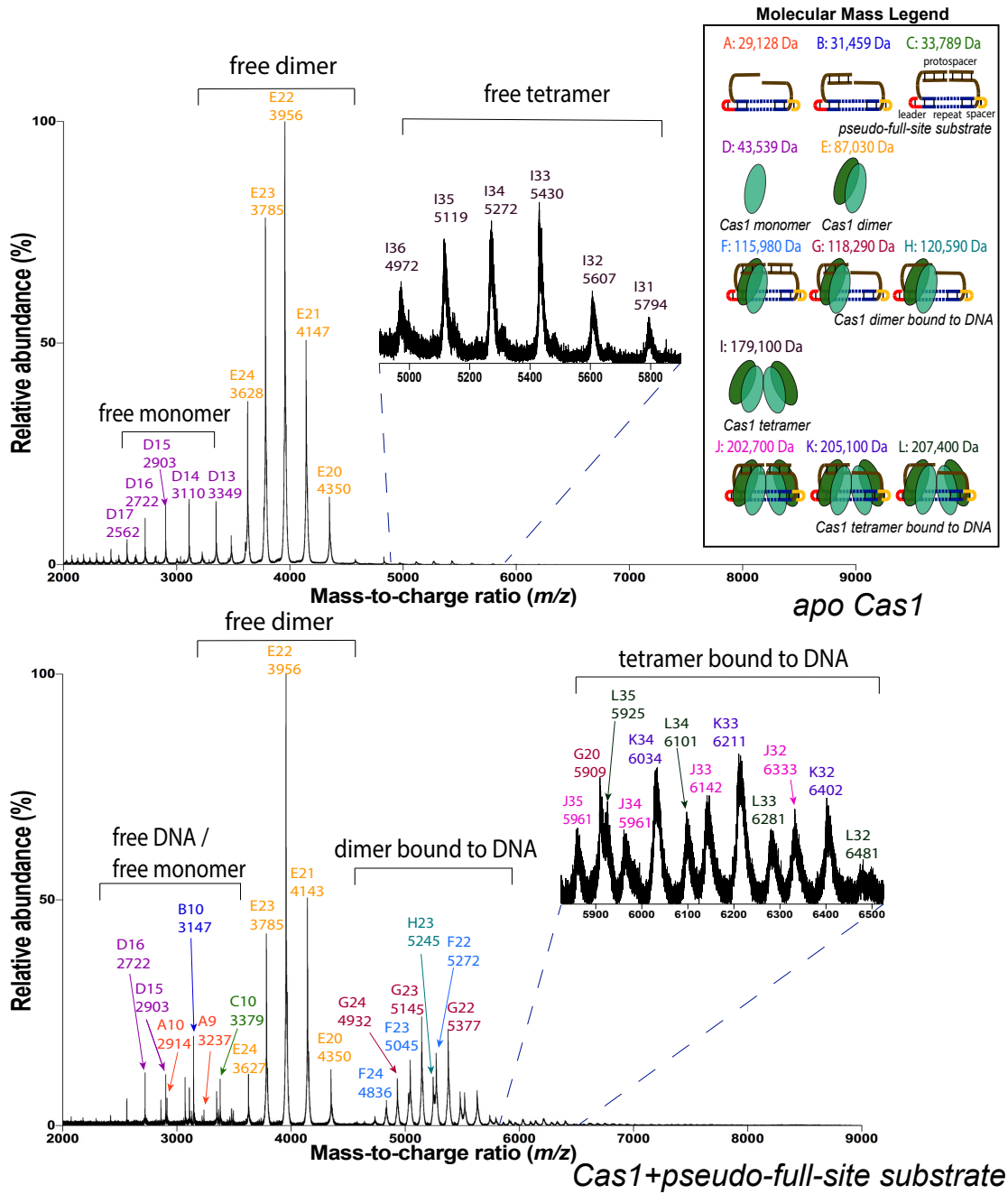J: 202,700 Da K: 205,100 Da L: 207,400 Da

Cas1 tetramer bound to DNA

Figure S6. Related to Figure 6. Native Mass Spectra of Apo V-C Cas1 Compared to V-C Cas1 Complexed with Pseudo-Full-Site Substrate

The molecular masses of the detected species are listed on the top right. The ions of different species are labeled with different colors. See Table S2 for nucleotide sequences.

Table S1. Related to Figure 1. IMG Genome Sequences of Type V-C Cas1 from Mouse Cecum and V-D Cas1 from Beetle Gut

Source	Identifier	Sequence
Passalidae beetle gut sample Sequenced at the JGI in 2012. Public data. PI: David Sherman	**Taxon ID**: 3300000114 **Scaffold ID**: IMNBL3_c10001072 **Gene ID**: IMNBL3_1000107225 **GOLD Analysis Project ID**: Ga0026078	MAHRIPLWLPYLSEVNVSASGVIHYVVYKGGEDSVRLKDISSIMIYGETDA SFSFKTFEKISRAGIPVIYTRRNIGSPMIIHGGLRYDAVDTVTAQILARE NKKKQCHIARQLLKAKMNSMKHLAPVFELPDNANINELRNIEAKHAKKYW REYYTYIGHPEWTRRGEETGKNPVSDTLNLTSKFICGIILRWTLYHHLSP SHGYFHVMTDYQSLVYDLIEPYRALFDIELIRLFKDNAHLEENWPGMAIN LMKNKLNKKCYCPLTRQVVTYHELTHGIVLSLKSYLIGRQRKFHIPLPGS PNGGRPSKVEFKLYARNAGKTDFWKVAKEVADLDFACLSVG*
Mouse cecum sample Sequenced at the JGI in 2016. Public data. PI: Michael Wagner	**Taxon ID**: 3300005460 **Scaffold ID**: Ga0073908_1000520 **Gene ID**: Ga0073908_1000520 11 **GOLD Analysis Project ID**: Ga0073908	MDQGNQTIENQTINCDQHPDFLWTWKSNKRGSRVSVWLPYFSQAKKIPRS KKWSVAYNGGSIEFDLKETDLIMFYGATGELPLEFLDDASKNGVMILIHR RNVLQPYVFYPSVIGDEEDILTKQIQFRTNERKRLYIAKTLIKKRLENMG STIPISAPLLRQLSAAKSIDEVRAIEANTTARYWNKWYENLNIETTRRKD HPINSALDAGSKFIYGVILRWLVFHRFSPNHGFMHQPTSYPSLVYDLMEP FRYMIENVCSAAWKRGERENSKIVALSLSFLKEELDKPCYVPATRQYVRK KNLLHGAVLALRSYLIGDMRKLVFPSEGVPNGGRPIKASYKLPGSMYDVG RKPPEIKQKDEICFDEVSQEESEE*

Table S2. Related to Figures 2-6, S1-S6. DNA Substrates Used in This Study

Description	Sequence	Figure
15mer protospacer	GCGCAGGCGGATAAG	S1
Fluorescent 15mer protospacer	(5' 6-FAM) GCGCAGGCGGATAAG	2
RC[1]	CTTATCCGCCTGCGC	2, S1
16mer protospacer	AGCGCAGGCGGATAAG	S1
Fluorescent 16mer protospacer	(5' 6-FAM) AGCGCAGGCGGATAAG	2
RC	CTTATCCGCCTGCGCT	2, S1
17mer protospacer	AAGCGCAGGCGGATAAG	S1
Fluorescent 17mer protospacer	(5' 6-FAM) AAGCGCAGGCGGATAAG	2
RC	CTTATCCGCCTGCGCTT	2, S1
18mer protospacer	AAAGCGCAGGCGGATAAG	S1
Fluorescent 18mer protospacer	(5' 6-FAM) AAAGCGCAGGCGGATAAG	2
RC	CTTATCCGCCTGCGCTTT	2, S1
18mer protospacer for sequencing (+) strand	CACGCTGGCCTCGGAGAG	3, S2, S3
RC ((−) strand)	CTCTCCGAGGCCAGCGTG	3, S2, S3
19mer protospacer	GAAAGCGCAGGCGGATAAG	S1
Fluorescent 19mer protospacer	(5' 6-FAM) GAAAGCGCAGGCGGATAAG	2
RC	CTTATCCGCCTGCGCTTTC	2, S1
20mer protospacer	AGAAAGCGCAGGCGGATAAG	S1
Fluorescent 20mer protospacer	(5' 6-FAM) AGAAAGCGCAGGCGGATAAG	2
RC	CTTATCCGCCTGCGCTTTCT	2, S1
21mer protospacer	AGAAAGTCGCAGGCGGATAAG	S1
Fluorescent 21mer protospacer	(5' 6-FAM) AGAAAGTCGCAGGCGGATAAG	2
RC	CTTATCCGCCTGCGACTTTCT	2, S1
22mer protospacer	AGAAAGTCGCAGGCTGGATAAG	S1
Fluorescent 22mer protospacer	(5' 6-FAM) AGAAAGTCGCAGGCTGGATAAG	2
RC	CTTATCCAGCCTGCGACTTTCT	2, S1
23mer protospacer	AGAAAGTCCGCAGGCTGGATAAG	S1
Fluorescent 23mer protospacer	(5' 6-FAM) AGAAAGTCCGCAGGCTGGATAAG	2
RC	CTTATCCAGCCTGCGGACTTTCT	2, S1
23mer protospacer with 3' P modification	AGAAAGTCCGCAGGCTGGATAAG (3'P)	S1
RC with 3' P modification	CTTATCCAGCCTGCGGACTTTCT (3'P)	S1
24mer protospacer	AGAAAGTCCGCAGGCATGGATAAG	S1

---

[1] RC: reverse complementary strand of the previous oligonucleotide

Fluorescent 24mer protospacer	(5' 6-FAM) AGAAAGTCCGCAGGCATGGATAAG	2
RC	CTTATCCATGCCTGCGGACTTTCT	2, S1
25mer protospacer	AGAAAGTCCGGCAGGCATGGATAAG	S1
Fluorescent 25mer protospacer	(5' 6-FAM) AGAAAGTCCGGCAGGCATGGATAAG	2
RC	CTTATCCATGCCTGCCGGACTTTCT	2, S1
33mer protospacer	GCGAGAATTACTACTCGTTCTGGTGTTTCTCGC	2
RC	GCGAGAAACACCAGAACGAGTAGTAATTCTCGC	2
1-nt overhang protospacer	AAGCGCAGGCGGATAAG	S1
RC	TTATCCGCCTGCGCTTT	S1
2-nt overhang protospacer (G)	AGCGCAGGCGGATAAG	4, 5, S1, S4
RC (T)	TATCCGCCTGCGCTTT	4, 5, S1, S4
RC (A)	TATCCGCCTGCGCTTA	S4
RC (C)	TATCCGCCTGCGCTTC	S4
RC (G)	TATCCGCCTGCGCTTG	S4
3-nt overhang protospacer	GCGCAGGCGGATAAG	S1
RC	ATCCGCCTGCGCTTT	S1
4-nt overhang protospacer	CGCAGGCGGATAAG	S1
RC	TCCGCCTGCGCTTT	S1
2-nt overhang protospacer (A)	AGCGCAGGCGGATAAA	S4
2-nt overhang protospacer (C)	AGCGCAGGCGGATAAC	S4
2-nt overhang protospacer (T)	AGCGCAGGCGGATAAT	S4
Pseudo-full-site half-protospacer - repeat - spacer - hairpin - spacer (pseudo-full-site substrate)	GGCGGCGGACCCATTATTGGAAGGGTTTATAAGGCTCACGAAGTGAG	5, 6, S5, S6
Pseudo-full-site half-protospacer - repeat - leader - hairpin - leader (pseudo-full-site substrate)	GGCGGCGGTCCTTATAAACCCTTCCAATAATGGGAGACTGAAAGTCT	5, 6, S5, S6
Pseudo-full-site half-protospacer	CCGCCGCC	5, 6, S5, S6
Half-site protospacer-repeat-spacer	ACGCTGGCCTCGGAGAGCCCATTATTGGAAGGGTTTATAAGGATCAA	5
Half-site protospacer	TCTCCGAGGCCAGCGTG	5
Half-site spacer-repeat-leader-hairpin-leader	TTGATCCTTATAAACCCTTCCAATAATGGGAGACTGAAAGTCT	5

CRISPR repeat target	AAGCAGGTCTTCTTTACTTTCCTTATAAACCCTTCCAATAATGGGAGACTAGCAAATATCTATTTGAAAGTCAAT	S3
RC	ATTGACTTTCAAATAGATATTTGCTAGTCTCCCATTATTGGAAGGGTTTATAAGGAAAGTAAAGAAGACCTGCTT	S3
lacZ hotspot target	CAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATG	S3
RC	CATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTG	S3
Sequencing Y-adapter strand 1[2]	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T	3, S2
Sequencing Y-adapter strand 2	GATCGGAAGAGCTGT	3, S2
(+)-strand protospacer-specific barcoding primer	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNWNNWNNCACGCTGGCCTCGGAGA*G	3, S2
(−)-strand protospacer-specific barcoding primer	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNWNNWNNCTCTCCGAGGCCAGCGT*G	3, S2